

УДК 811.111'01/01

О. Ю. Андрушенко,

кандидат філологічних наук, доцент

(Житомирський державний університет імені Івана Франка)

ФОРМАЛЬНИЙ АНАЛІЗ КОНТЕКСТУ У ДІАХРОННИХ ДОСЛІДЖЕННЯХ ІНФОРМАЦІЙНОЇ СТРУКТУРИ РЕЧЕННЯ

У статті розглядаються сучасні підходи до дослідження інформаційної структури (ІС) речення в історичних студіях. Грунтуючись на надбаннях генеративної граматики, корпусної та комп'ютерної лінгвістики останніх років, автор пропонує власну методику вивчення інформаційно-структурних компонентів, яка полягає у комплексному застосуванні генеративних моделей ІС та формальних методів аналізу контексту, перспективним є побудова дериваційної моделі речення, дерева залежностей та концептуального графа речення.

Інформаційна структура (ІС) речення відображає його членування на відомий зміст (топік) та нове, невідоме (фокус). Вважається, що фокус – це гіпотетично створений компонент, який має семантико-прагматичний корелят (нова інформація), синтаксичний корелят (фокус-ознака) та фонологічний корелят (головний наголос речення). Аналогічні відповідники притаманні і топіку.

Складність аналізу інформаційної структури речення пов'язана із її контекстною залежністю, яка, у свою чергу, іноді знаходиться у позамовній площині. Актуальним завданням є визначення в реченні "меж" топіку та фокусу, що завжди вважалося наріжним каменем теорії ІС. Нового поштовху у вирішенні цих питань надали останні досягнення генеративної теорії (Програма мінімалізму Н. Хомського) [1] із залученням здобутків корпусної та комп'ютерної лінгвістики [2; 3]. Не залишився поза увагою й діяхронний аспект означеної проблеми [4]. Проте вивчення ІС речення в історичній перспективі пов'язане із певними складнощами: обмеженою кількістю писемних пам'яток, їх різноманітністю, відсутністю носіїв давніх мов. Останнє унеможливило вивчення в реченні просодичних явищ, які в сучасних мовах широко використовуються при визначенні фокусу та топіка.

Мета нашого дослідження – довести доцільність застосування при вивченні давніх пам'яток формальних методів контекстуального аналізу, які у сукупності із новітніми генеративними синтаксичними моделями ІС речення забезпечують достовірне визначення ІС компонентів.

Згідно з найбільш відомою генеративною моделлю Л. Ріцці [5], компоненти ІС речення пов'язані між собою формулою:

(1) $[ForceP\ Force[_{TopP}\ Top[_{FocP}\ Foc[_{TopP}\ Top[_{FinP}\ Fin[IP\ \dots]]]]]]$, де ForceP – силова фраза, встановлює відношення речення до контексту; TopP (топік) презентує стару інформацію в контексті (* позначає, що проєкції Top можуть повторюватися); Foc (фокус) експлікує нову інформацію; FinP (фінітна група) містить такі елементи як час, спосіб, стан тощо; IP – проєкції, що слідує нижче у деривації.

Здобутком цієї концепції є формальний аналіз речення, проте контекст в ній враховано мінімально, що не дозволяє вивчити ІС у повному обсязі.

Спробу вдосконалити модель Л. Ріцці здійснила П. Бенінка [6]. Так, ІС речення представлена наступним чином:

(2) $[DiscP[ForceP[TopP[FocP[FinP[IP\ \dots]]]]]]$, де DiscP – дискурсивна фраза, що забезпечує місце для "основогенеруючого" тематичного елемента (винесений топік).

На наш погляд, цій концепції притаманні ті самі недоліки, що і моделі Л. Ріцці, тобто слабкий зв'язок із контекстом. Окрім того, у структурі П. Бенінки наявний лише один топік, а також відсутня визначена позиція TopP по відношенню до FocP.

Останнім часом науковці дедалі частіше вказують на взаємозв'язок між ІС речення та дискурсом, тобто на необхідність залучення додаткової інформації із дискурсу. У термінах М. Кріфки [7] цей процес отримав назву "управління спільним змістом", який виходить за межі досліджуваного речення. Специфіка діяхронних розвідок спонукає нас при вивченні інформаційно-структурних явищ звернутися до контексту давніх пам'яток. Цей підхід вимагає зробити низку припущень. Зокрема зазначимо, що дослідники давніх текстів завжди знаходяться на позиції слухача. Припускаємо також, що тексті пам'ятки існує загальновідома інформація, яка згідно теорії ІС, є старою та корелює із топіком речення. Наприклад, в *The York-Toronto-Helsinki Parsed Corpus of Old English Prose (YCOE)* ця інформація неодноразово згадується в текстах, що за хронологією є старшими за писемну пам'ятку, до якої належить аналізоване речення. Окрім того, необхідно визначитись, яка частина тексту пам'ятки слугуватиме матеріалом для контекстного аналізу при визначенні топіка і фокусу речення. Відомо, що контекстом є та частина писемного твору, яка дає змогу точно встановити значення окремого слова чи виразу, що входить до її складу. Зробимо припущення, що такими частинами у нашому випадку є окремі епізоди давнього тексту. Формальний пошук таких фрагментів можна здійснити, спираючись на праці У. Ленкер [8: 227-231]. Так, дослідниця стверджує, що давньоанглійські лексеми *witodlice*, *soplice* є маркерами меж

епізоду, оскільки вони доволі часто зустрічаються на початку параграфа, та у манускриптах позначені спеціально прикрашеними початковими літерами.

Суть запропонованої методики полягає у наступному. Вивчення інформаційних характеристик речення будемо здійснювати із використанням давньоанглійських текстів із корпусу *YCOE*. Виокремленні у випадковій послідовності речення аналізуємо спочатку згідно однієї із генеративних синтаксичних моделей, напр. з моделлю Л.Ріцці (див. формулу 1):

(3) *Se geworhte micelne dom on ðæm gefeohte* (Oros). – He gained great glory in that battle.

Представимо ІС речення (3) у вигляді дериваційної моделі:

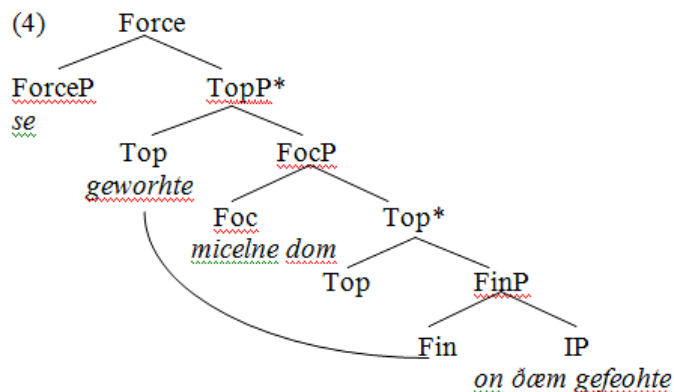


Рис. 1 Дериваційна модель інформаційної структури речення.

Отже, генеративна модель ІС речення дозволила визначити його фокус і топик. Вважатимемо ці результати попередніми, та для їх остаточного підтвердження проаналізуємо контекст речення. У комп'ютерній та корпусній лінгвістиці для завдань пошуку інформації широко застосовують дерева залежностей [2; 3]. Речення представлено графом, вершини якого відповідають словам, а дуги – синтаксичному зв'язку, що їх поєднує. Оскільки на Рис. 1 також зображений граф, наше речення досить легко трансформується у дерево залежностей:

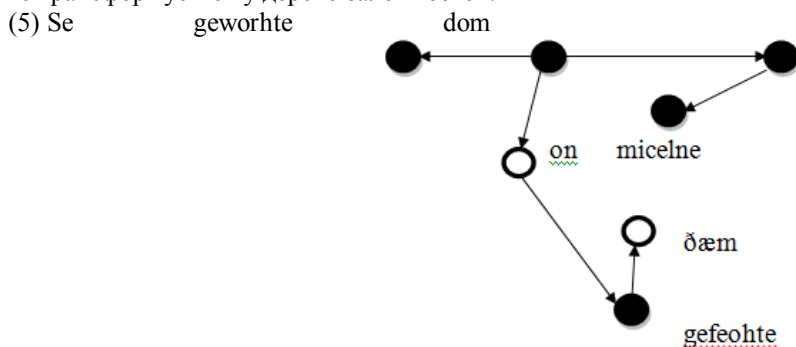


Рис. 2 Дерево залежностей речення.

Звернемо увагу, що вершини дерева залежностей не є рівноправними. Одні з них експлікують повнозначні слова (*geworhte*, *dom*), інші – "стоп-слова": прийменники, союзи тощо (*on*, *ðæm*). Ці вершини будуть виключені із подальшого аналізу контексту, оскільки їх змістове навантаження мінімальне. Існують спеціальні комп'ютерні програми (технологія XML для аналізу текстової інформації), які автоматично будують граф залежностей та виконують пошукові операції [3].

Таким чином, виокремлюючи фрагменти графа, що відповідають топіку та фокусу речення, досить легко здійснити формальний пошук старої / нової інформації як у межах епізоду писемної пам'ятки, так і в усьому корпусі текстів (перевірка, чи не є аналізована інформація загальновідомою).

Відмітимо, що наведений вище метод аналізу контексту із використанням графа залежностей є доволі спрощеною процедурою. У реальних текстах у межах епізоду лексемі "*ladteow*" у попередніх реченнях може відповідати "*ealdor*", "*hēafordman*", "*here-toga*" тощо. Отже, виникає потреба при аналізі контексту для визначення ІС характеристик речення використовувати інший граф – концептуальний. Концептуальний граф як семантична модель речення дозволяє знайти у тексті словосполучення у вигляді пар концептів, пов'язаних певними відносинами, наприклад:

(6) *þaet Darius hie mid gefeohte secan wolde, hie 25 acuron endlefan pusend monna, 7 him ongean foran, 7 pone cyning set paere dune metton pe mon haett Morotthonie 2. Heora ladteow waes haten Htesseus; se waes mid his daedum snelra ponne he maegenes haefde ; se geworhte micelne dom on ðaem gefeohte. þa wearp tu hund pusenda 30 Persea ofslaegen, 7 pa opre gefliemed. þa he eft haefde fird gegaderod on Perseum, 7 paet wrecan þohte, þa gefor he (Oros.)*

Так, лексемі *ladteow* (*general, leader*) далі по оповіді відповідає власна назва *Htesseus*, вказівний займенник *se*, та особовий займенник *he*. Це можна визначити лише з самого контексту. Тому вказівний займенник в аналізованому реченні відноситься до лексеми *ladteow* "полководець". Також з самого контексту відомо, що *ladteow* стосується полководця короля Дарія. Проте, ця сама лексема далі по тексту може позначати іншого полководця, оскільки у самому творі описуються безліч воєнних подій.

Концептуальний граф будується автоматично (технологія аналізу текстових даних Text Mining), записи (наприклад, частина речення, що експлікує фокус) йдуть у довільному порядку, стоп-слова не аналізуються [2].

На відміну від графа залежностей, концептуальний граф містить два типи вершин: концепти та концептуальні відносини. Напр.:

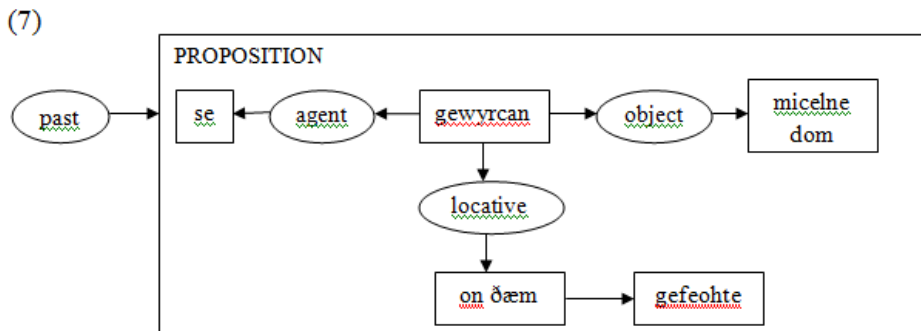


Рис. 3 Концептуальний граф речення.

Представлення концептуального графа у притаманній комп'ютерній лінгвістиці текстовій формі має такий вигляд:

(8) (PAST) -> [[SE] <- (AGENT) <- [GEWYRCAN] -> (OBJECT) -> [MICELNE DOM]]
 |
 (LOCATIVE)
 [ON ÐÆM GEFEONTE]

Використовуючи методи пошуку старої / нової інформації із використанням концептуальних графів, дослідник давніх текстів має можливість впритул наблизитись до найбільш точного визначення інформаційно-структурних компонентів та їх "меж" у реченні. Зазначимо, що новітні розробки у царині комп'ютерної та корпусної лінгвістики пов'язані із створенням семантичних мереж – сукупності заданої множини концептуальних графів із описом їх взаємозв'язків та зануренням у контекст певної сфери пошуку. Такою сферою можуть бути й давні писемні пам'ятки.

Діахронні розвідки ІС речення є вкрай важливими. К. Ламбрехт зазначає [9: 28-29], що саме вони здатні пролити світло на те, яким чином розвиваються морфологія, просодія, семантика і прагматика, та як вони взаємодіють. Оскільки можливості дослідників давніх пам'яток вкрай обмежені, слід застосовувати весь комплекс методів та підходів (синтаксичних, семантико-прагматичних, структурних) для досягнення результату. На наш погляд, запропонована методика вивчення ІС речення в історичній площині, яка полягає у комплексному використанні генеративних синтаксичних моделей та методів формального аналізу контексту, спроможна забезпечити такий результат.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ ТА ЛІТЕРАТУРИ

1. Chomsky N. Deep Structure, Surface Structure and Semantic Interpretation / Noam Chomsky // *Semantics : an Interdisciplinary Reader in Philosophy, Linguistics, and Psychology*. – Cambridge: CUP, 1995. – P. 183–216.
2. Conceptual Structures for Knowledge Creation and Communication // Eds. Moor A., Lex W., Ganter B. – Berlin ; Heidelberg: Springer, 2004. – 405 p.
3. Kübler S., McDonald R., Nivre J. Dependency Parsing / Eds. S. Kübler, R. McDonald, J. Nivre . – San Francisco : Morgan & Claypool, 2009. – 115 p.
4. Kemenade A. van. Discourse Relations and Word Order Change / Ans van Kemenade // *Information Structure and Word Order Change : New Approaches to Word Order Variation in Germanic*. – Berlin; N.-Y. : Mouton der Gruyter, 2009. – P. 91–121.
5. Rizzi L. The Fine Structure of the Left Periphery / Luigi Rizzi // *Elements of Grammar*. – Dordrecht : Kluwer, 1997. – P. 281–337.
6. Benincá P. The Position of Topic and Focus in the Left Periphery / Paola Benincá // *Current Studies in Italian Syntax. Essays Offered to Lorenzo Renzi*. – Amsterdam : Elsevier-North Holland, 2001. – P. 39–64.
7. Krifka M. Basic Notions of Information Structure / Manfred Krifka // *Interdisciplinary Studies of Information Structure 6*. / Eds. Fery C. ; Krifka M. – Potsdam : University of Potsdam, 2007. – P. 13–55.
8. Lenker U. Argument and Rhetoric. Adverbial Connectors in the History of English / Ursula Lenker. – Berlin ; N.-Y. : Walter de Gruyter GmbH & Co., 2010. – 293 p.

9. Lambrecht K. Information Structure and Sentence Form : Topic, Focus and the Mental Representations of Discourse Referents / Knud Lambrecht. – Cambridge : CUP, 2000. – 388 p.

REFERENCES (TRANSLATED & TRANSLITERATED)

1. Chomsky N. Deep Structure, Surface Structure and Semantic Interpretation / Noam Chomsky // Semantics: an Interdisciplinary Reader in Philosophy, Linguistics, and Psychology. – Cambridge: CUP, 1995. – P. 183–216.
2. Conceptual Structures for Knowledge Creation and Communication // Eds. Moor A., Lex W., Ganter B. – Berlin; Heidelberg: Springer, 2004. – 405 p.
3. Kübler S., McDonald R., Nivre J. Dependency Parsing / Eds. S. Kübler, R. McDonald, J. Nivre. – San Francisco: Morgan & Claypool, 2009. – 115 p.
4. Kemenade A. van. Discourse Relations and Word Order Change / Ans van Kemenade // Information Structure and Word Order Change: New Approaches to Word Order Variation in Germanic. – Berlin; N.-Y.: Mouton der Gruyter, 2009. – P. 91–121.
5. Rizzi L. The Fine Structure of the Left Periphery / Luigi Rizzi // Elements of Grammar. – Dordrecht: Kluwer, 1997. – P. 281–337.
6. Benincá P. The Position of Topic and Focus in the Left Periphery / Paola Benincá // Current Studies in Italian Syntax. Essays Offered to Lorenzo Renzi. – Amsterdam: Elsevier-North Holland, 2001. – P. 39–64.
7. Krifka M. Basic Notions of Information Structure / Manfred Krifka // Interdisciplinary Studies of Information Structure 6. / Eds. Fery C.; Krifka M. – Potsdam: University of Potsdam, 2007. – P. 13–55.
8. Lenker U. Argument and Rhetoric. Adverbial Connectors in the History of English / Ursula Lenker. – Berlin; N.-Y.: Walter de Gruyter GmbH & Co., 2010. – 293 p.
9. Lambrecht K. Information Structure and Sentence Form: Topic, Focus and the Mental Representations of Discourse Referents / Knud Lambrecht. – Cambridge: CUP, 2000. – 388 p.

Матеріал надійшов до редакції 13.05. 2013 р.

Андрушенко Е. Ю. Формальный анализ контекста в диахронических исследованиях информационной структуры предложения.

В статье рассматриваются современные подходы к исследованию информационной структуры (ИС) предложения в исторических студиях. Основываясь на достижениях генеративной грамматики, корпусной и компьютерной лингвистики последних лет, автор предлагает собственную методику изучения информационно-структурных компонентов, заключающуюся в комплексном применении генеративных моделей ИС и формальных методов анализа контекста, перспективным является построение деривационной модели предложения, дерева зависимостей и концептуального графа предложения.

Andrushenko O. Yu. Context Formal Analysis of Sentence Information Structure in Diachronic Studies.

The article highlights modern approaches to sentence information structure (IS) investigation in historical studies. On the basis of recent advances in generative grammar, corpus and computer linguistics, the author proposes original methodology of investigating information-structural components, which presupposes complex implementation of IS generative models and formal methods of context analysis. Potential in this respect is building sentence derivational model, dependency and conceptual graphs.